

PORTFOLIO CHOICE AND ESTIMATION RISK.
A COMPARISON OF BAYESIAN TO HEURISTIC APPROACHES*

BY

ULF HEROLD AND RAIMOND MAURER

ABSTRACT

Estimation risk is known to have a huge impact on mean/variance optimized portfolios, which is one of the primary reasons to make standard Markowitz optimization unfeasible in practice. This issue has attracted new interest in the last years, and several approaches to incorporate estimation risk into portfolio selection have been developed only recently. In this article, we review these approaches as well as some older ones and compare them in an empirical out-of-sample study. The approaches can be classified along two criteria. First, we can differentiate heuristic approaches (restricting portfolio weights and employing simulation techniques) and those based on Bayesian statistics (shrinking the portfolios towards a pre-determined target). Second, the assumptions about the return-generating process differ, either assuming returns to be IID distributed or to be partly predictable. The central result of our empirical study is that all of the IID approaches, whether they account for estimation risk or not, are not superior to simple investment strategies like holding the market portfolio. A risk-adjusted outperformance is possible only if sample means are substituted with conditional expected return estimates. Furthermore, the Bayesian approaches reduce turnover and stabilize portfolio weights.

KEYWORDS

Portfolio selection, estimation risk, predictive regressions.

I. INTRODUCTION

Estimation risk is known to have a huge impact on *Markowitz* (1959) mean/variance (MV) optimized portfolios. It leads to unstable and extreme portfolio

* The authors profited from the comments of two anonymous referees. The authors also thank Thanh H. Vo for excellent research assistance. This research project was supported by the German Investment and Asset Management Association and the E-finance lab at the Goethe-University Frankfurt.

weights over time and along portfolios on the MV efficient frontier. MV optimized portfolios lack of diversification and show poor out-of-sample performance. Hence, estimation risk is one of the primary reasons to make standard MV optimization unfeasible in practice. *Michaud* (1998, p. xiv) summarizes this fact using the term “Markowitz optimization enigma” and, in a paper reviewing the dialogue between theory and practice in financial economics, *Banz* (1997, p. 389) concludes, “I believe that (...) estimation risk is one of the great neglected areas of modern finance”. This issue has attracted new interest in the last years, and several approaches to incorporate estimation risk into portfolio selection have been developed only recently. The approaches can be classified along two criteria. First, we can differentiate heuristic approaches and those based on Bayesian statistics. Second, the assumptions about the return-generating process differ, either assuming returns to be independently and identically distributed (IID) or to be partly predictable.

The objective of the heuristic approaches is to produce more diversified portfolios. This is achieved either by introducing constraints on portfolio weights or by performing simulations and averaging over the MV efficient portfolios obtained in these simulation trails. On the other hand, Bayesian statistics is used to quantify estimation errors and incorporate them into the portfolio selection process. The idea of Bayesian inference is to combine extra-sample, or prior, information with sample data. The portfolio is shrunk towards a pre-determined shrinkage target, depending on the degree of noise in the sample. The most popular approach is the Bayes/Stein estimation procedure developed by *Jorion* (1985, 1986) which employs a statistically motivated prior. He sets expected returns to a common value in the prior and shrinks towards the minimum-variance portfolio. In contrast, *Pastor* (2000), *Pastor/Stambaugh* (1999, 2000) and *Wang* (2001) start with an economic prior. They suggest to incorporate asset-pricing models into portfolio selection and shrink towards the optimal portfolios implied by that asset-pricing model.

These approaches operate under the classical assumption of IID returns (“unconditional strategies”). In investment practice and in modern financial economics, this classical assumption about the return-generating process is called into question. Instead, expected returns are assumed to be time-varying in a predictable fashion. Instrumental variables like dividend yields or term spreads are employed to predict expected returns (“conditional strategies”). However, the estimation risk issue is still a severe problem, as the parameters of the predictive regressions or other types of forecasting models need to be estimated from the data. A possibility to account for the parameter uncertainty that is attached to the coefficients of the forecasting model is to apply Bayesian regression models and shrink the optimal portfolio, based on the conditional expected return estimates, again either towards the minimum-variance or the market portfolio.

The objective of this paper is to review the approaches mentioned above and compare them in an empirical out-of-sample study, using a long history of U.S. data and a shorter history of EMU stock returns. The existing literature that

compares some of the approaches reviewed above comes up with different results. *Chopra et al.* (1993) and *Jorion* (1985, 1991) find the Bayes/Stein procedure superior to standard MV analysis. In contrast, *Fletcher* (1997) and *Grauer/Hakansson* (1995) cannot confirm the better performance of shrinkage estimators. *Michaud* (1998) shows in a Monte Carlo simulation experiment that simulation-based approaches (called “resampled efficiency”) slightly increases the Sharpe ratio compared to standard MV optimization. *Fletcher/Hillier* (2001) cannot find that resampling systematically enhances out-of-sample performance. Neither of these studies applies the Bayesian approach of shrinking towards the market portfolio. What differentiates our research further from this literature is that we do not evaluate performance figures only but also examine turnover and portfolio weight statistics to obtain information whether the approaches are feasible in investment practice.

Our article is closely related to a recent paper by *DeMiguel/Garlappi/Uppal* (2005). They investigate a similar set of unconditional strategies and perform an extensive empirical study with a variety of data sets. The main distinction to our analysis is the choice of portfolio selection models in the context of time-varying expected returns. *DeMiguel et al.* examine two models outlined in the book of *Campbell/Viceira* (2002) that employ multi-period dynamic optimization and incorporate hedging demands. These models have been originally designed for investors with a long horizon – in fact, the two models assume that investors live infinitely long. *DeMiguel et al.* include these models in their backtest with monthly data and evaluate the monthly returns produced by these models. So they apply these models for short-term portfolio management.

In contrast, we investigate conditional strategies based on predictive regressions. We have chosen this model because a backtest is commonly set up in a “myopic” way: The investor has an investment horizon of one period (one month), and the performance results of the strategies are consequentially compared on the basis of one-month returns. The predictive regressions estimate the expected return for the next period and hence operate with an investment horizon of one month. This enables a fair comparison to the unconditional strategies that are also based on one-month horizons. Furthermore, we incorporate estimation risk via Bayesian regression models. In contrast, *DeMiguel et al.* do not account for estimation risk in their conditional models. Two minor differences are that we have chosen a broader set of state variables and assets for the conditional models and that we check for statistically significant differences in the Sharpe ratios with the more general GMM test while *DeMiguel et al.* employ the classical *Jobson/Korkie* (1981a) test procedure.

The central result of our empirical study is that all of the approaches that operate under the IID setting, whether they account for estimation risk or not, are not superior to simple investment strategies like holding the market portfolio, the equally-weighted portfolio or the minimum-variance portfolio which refrain from estimating expected returns at all. In this regard, *DeMiguel/Garlappi/Uppal* (2005) come to the same conclusion. We find that a risk-adjusted

outperformance is possible only if sample means are substituted with conditional expected return estimates using exogenous information. This is “good news” for investment practice and encourages active managers to stay on the search for valuable exogenous information. Here, the findings of *DeMiguel et al.* are disappointing: their conditional strategies do not turn out to add value so that the recommendation for the investor is to naively diversify (assign equal weights). Another result of our empirical study is that in any case – for conditional and unconditional strategies –, the Bayesian approaches reduce turnover and stabilize portfolio weights.

The paper is organized as follows. Section II briefly summarizes the impact of estimation risk on MV optimized portfolios. Sections III to V contain a description of each of the approaches we review. Section VI gives the results of the various out-of-sample studies, and the last Section concludes.

II. IMPACT OF ESTIMATION RISK ON MV OPTIMIZED PORTFOLIOS

Markowitz (1959) mean/variance (MV) efficiency is the classic paradigm of modern finance for allocating capital among risky assets. Markowitz shows how to construct efficient portfolios. The MV objective function is given by

$$\omega' \mu - \frac{\lambda}{2} \omega' \Sigma \omega, \quad [1]$$

where ω is the $N \times 1$ vector of portfolio weights, μ is the $N \times 1$ vector of expected returns, Σ is the $N \times N$ covariance matrix of returns, and λ denotes risk aversion. In each period, the investor trades off expected portfolio return, $\omega' \mu$, versus portfolio variance, $\omega' \Sigma \omega$. He chooses his portfolio ω to maximize the value of the objective function given in [1]. The minimum variance frontier comprises all portfolios that have minimum variance for a given level of expected return. The MV efficient frontier is the upward sloping portion of the minimum variance frontier.

Inputs are expected mean future returns for each asset, expected volatility of returns around the future expected means and the matrix of expected correlations of all returns. The optimization algorithm takes these inputs as parameters of known probability distributions. However, in reality, they are estimates of parameters of unknown probability distributions. Even if expected returns, variances, and correlations were known with certainty, MV optimized portfolios would not beat all other portfolios in every future investment period, because return realizations will differ from their expected values (intrinsic risk). Markowitz portfolios will be optimal on average. Estimating the unknown parameters involves an additional source of risk: estimation error, or estimation risk. So an asset's total risk is composed of two components: intrinsic risk and estimation risk.

The impact of estimation risk on optimal portfolios has been explored in the previous literature. *Chopra/Ziembra* (1993) find that errors in means are

about ten times as important as errors in variances, and errors in variances are about twice as important as errors in covariances. *Best/Grauer* (1991) show that optimal portfolios are very sensitive to the level of expected returns. They note that “a surprisingly small increase in the mean of just one asset drives half the securities from the portfolio. Yet the portfolio’s expected return and standard deviation are virtually unchanged” (p. 325).

When based on sample means and covariances, MV optimized portfolios regularly exhibit a low degree of diversification. Only few assets are included in the optimal portfolio. They show sudden shifts in allocations along the efficient frontier and are also very unstable across time. As *Michaud* (1989) pointed out, these unintuitive and extreme solutions are a consequence of optimizers being “estimation error maximizers”. MV optimizers overweight those assets that have large estimated expected returns, low estimated variances and low estimated correlations to other assets. These assets are the ones most likely to have large estimation errors. Consequently, *Jorion* (1985), using rolling-window estimates based on actual data, and *Jobson/Korkie* (1981b), employing a simulation approach, document poor out-of-sample performance of MV optimized portfolios compared to non-optimized, heuristic approaches (equally-weighted portfolio and market portfolio).

III. HEURISTIC APPROACHES TO INCORPORATE ESTIMATION RISK

1. Restricting portfolio weights

In investment practice, MV analysis is performed – if at all – only after including a variety of weight restrictions. Some constraints must be taken into account due to legal considerations (e.g., short-selling restrictions). But in their internal guidelines, asset managers usually impose an extensive set of additional constraints. By construction, constraints enforce diversification, which in turn is expected to improve out-of-sample performance. Using a simulation setup, *Frost/Savarino* (1998) show that short-selling and upper weight restrictions lead to an enhanced performance. *Eichhorn et al.* (1998) confirm this result and furthermore come to the conclusion that constraints reduce volatility and short-fall risk. In an out-of-sample study, *Grauer/Shen* (2000) find realized volatility to decline, but only at the cost of realized return.

From a theoretical point of view, the question arises whether MV efficient portfolios should be diversified at all. *Green/Hollifield* (1992) started this debate by noting that even the minimum-variance portfolio (MVP), constructed from individual stocks without constraints, usually exhibits extreme long/short positions. It should not be strongly affected by estimation errors as it does not require any expected returns. *Green/Hollifield* (1992) argue that stock returns exhibit a factor structure which is the reason for the extreme weights of the MVP. When applying principal component analysis to the sample covariance matrix, a dominant factor is found. This factor probably also exists in the structure of

the population returns. Hence, even if we knew the true parameters, the MVP will consist of extreme positions. They conclude that MV efficient portfolios will not be diversified and that investors should invest in these portfolios to achieve a significant risk reduction.¹

Then, placing restrictions on portfolio weights is a mis-specification. However, the covariance matrix is also subject to estimation errors, and the constrained MVP often performs better than its unconstrained counterpart. *Jagannathan/Ma* (2003) conduct an out-of-sample study to investigate which of both effects – mis-specification by imposing the “wrong constraints” or estimation errors – is predominant. They come to the conclusion that the impact of estimation errors is the more important one and hence, restrictions are justified.

Therefore, restrictions are an adequate means to reduce the impact of estimation risk on MV efficient portfolios. The crucial point, of course, is how to design them. Constraints must not be arbitrary. If constraints are set very tight, they predetermine the optimal portfolio to a large extent, and the problem is just moved to a different level. How to determine constraints is not a central issue of this paper but at least, we want to point out to a promising solution which has been developed by *Grinold/Easton* (1998) and which is also used by *Jagannathan/Ma* (2003). A useful way to assess the impact of the various constraints one imposes is to recognize that mathematically, each constraint is multiplied with a Lagrange multiplier and then enters the MV objective function [1]. The output of an optimization algorithm consists of the optimal weights and, in addition, the value of the Lagrange multipliers. If a constraint is binding, its multiplier is different from zero.

Starting with the initial input variables (expected returns and covariance matrix) which lead to an optimal portfolio under several constraints, the Lagrange multipliers can be used to obtain a modified set of input variables which yield the same portfolio in an optimization without any constraints. Therefore, introducing constraints implicitly adjusts the expected returns and covariances. Now, we can evaluate which constraints are most binding and which expected returns (or covariances) are adjusted due to which constraint. In a next step, we can ask whether we should relax some of the constraints or whether our input variables actually are reliable.² For those who want to work with constraints (e.g., in order to control their portfolio managers), this procedure gives some guidelines on how to design them. The others can tackle the

¹ To construct the MVP, *Green/Hollifield* (1992) first group the stocks along their betas. These groups represent portfolios with a low level of specific risk due to diversification. Then they combine these sub-portfolios so that the resulting portfolio shows a beta coefficient of zero. This portfolio has no systematic and no specific risk but it exhibits extreme long/short positions because it is short in high-beta-stocks and long in low-beta-stocks.

² In fact, this procedure only works when expected returns are given in a quantitative form. In practice, portfolio managers often only form market views or qualitative judgements. *Herold* (2003) provides a Bayesian approach to translate this market views into optimal portfolios. This approach can also be employed to design restrictions.

problem directly – by explicitly adjusting the input variables. This is what the Bayesian approaches amount to do.

2. Bootstrapping

A classical approach to account for estimation risk is the bootstrapping procedure by *Efron* (1979). Applied to portfolio optimization, this means to draw repeatedly from the empirical return distribution, compute efficient frontier portfolios based on these resampled returns, and then average over the portfolio weights. This way the portfolios show desirable attributes: they exhibit a higher degree of diversification, and their composition is less prone if a new observation is added to the data set.

The idea to use resampling methods in the context of portfolio selection and estimation risk goes back to *Jobson/Korkie* (1981b). *Jorion* (1992) also employs this simulation technique to show the variability that is implicit in the weights of efficient portfolios based on sample estimates. Resampling gives the full distribution of portfolio weights and not just a point estimate as MV analysis does. A special case is to resample from a multivariate normal distribution (instead of the empirical return distribution). This is the suggestion of *Michaud* (1998). We follow the more general bootstrap approach. For many data sets (where stock index returns are approximately normally distributed), the results are almost identical. Theoretical objections against the resampling procedure are raised by *Scherer* (2002).

The bootstrapping procedure can be summarized as follows: First, we compute the sample mean vector and the sample covariance matrix, $\hat{\mu}$ and $\hat{\Sigma}$, from T observations, and calculate the MV efficient portfolio corresponding to a given risk aversion coefficient, λ , in the MV utility function [1]. Then we draw T monthly returns from the data (with replacement). The returns are drawn across the N asset classes to preserve the historical correlation structure. This way, we obtain a new set of optimization inputs (μ, Σ) from this statistically equivalent sample, and again calculate the MV portfolio corresponding to λ . We repeat this step many times (we do it 500 times). Finally, we average over the weights of all simulated portfolios. This averaging procedure ensures that the weights still sum up to unity.

Resampling enforces diversification. This can be illustrated with the maximum return portfolio, i.e. the portfolio with $\lambda = 0$. In classical MV optimization, the maximum return portfolio consists solely of one asset class (under short-selling constraints). With resampling, it is not always this asset class that exhibits the highest mean return in a simulation trail. In each path, the maximum return portfolio consists of one asset class, but as this asset class varies in the simulations, averaging over weights of the simulations will bring up a maximum return portfolio which is exposed to several asset classes. There are also less sudden shifts in portfolio weights along the bootstrapped efficient frontier.

IV. BAYESIAN APPROACHES UNDER THE IID SETTING

1. Bayesian setup

In Bayesian statistics, some prior information is introduced and combined with the sample. More formally, let $y' = (y_1, y_2, \dots, y_T)'$ be a sample of T identically and independently distributed observations from an unknown probability density function (pdf), $p(y|\theta)$, where θ is the $K \times 1$ vector of parameters to be estimated. The classical statistical perspective assumes that there exists some true value of θ . This true value is unknown but a fixed number. Using, e.g., Maximum Likelihood techniques, an estimator $\hat{\theta}$ is constructed from the sample observations, which maximizes the sample likelihood. In contrast, Bayesian statistics treats θ as a random variable. All information that is known about θ before drawing the sample is summarized into the prior pdf $p(\theta)$. The posterior pdf combines the prior pdf and the sample and is given by:

$$p(y|\theta) \propto p(\theta)p(y|\theta), \quad [2]$$

where $p(y|\theta)$ is the likelihood function. From there, the predictive pdf of a future observation, \bar{y} , is obtained. It is the distribution of \bar{y} conditional on the observed data y and it is derived from the likelihood function of a future observation and the posterior distribution by integrating over θ :

$$p(\bar{y}|y) = \int p(\bar{y}|\theta)p(\theta|y)d\theta. \quad [3]$$

The idea to use Bayesian inference to incorporate estimation risk into portfolio selection goes back to *Barry* (1974) and *Klein/Bawa* (1976). They impose a diffuse or non-informative prior on the parameters $\theta = (\mu, \Sigma)$. The predictive pdf is then given by the sample mean vector and by the sample covariance matrix multiplied by a constant factor. Therefore, the compositions of the efficient frontier portfolios do not change. However, the investor will choose a different portfolio, one with less risk, i.e. he will move to the left on the efficient frontier. As the impact of a diffuse prior is rather small and as it leaves the estimator for the expected returns unchanged, it is not widely used. The Bayesian approaches we review impose an informative prior. In this case, the sample means are shrunk to the prior values, depending on the degree of estimation error in the sample and on the precision of the prior.

2. Shrinking towards the minimum-variance portfolio

As estimates of expected returns are prone to estimation errors, *Jorion* (1985, 1986) shrinks the MV efficient portfolios towards the MVP. The MVP is less vulnerable to estimation risk as it does not make use of any information about expected returns. The rationale is that we can estimate the covariance matrix

from the sample returns quite precisely, while estimation errors in the means are tremendous (although there are still estimation errors involved when constructing the MVP, as we have discussed above). *Jorion* sets the prior means to a common value across all N assets. He specifies the prior as

$$\mu \sim N(1_N \mu_0, \Sigma / \varphi), \quad [4]$$

where μ_0 is the expected return of the MVP and φ determines the prior precision. Assuming returns to be multivariate normally distributed and assuming the covariance matrix to be known, combining the prior with the sample likelihood function yields the posterior distribution with mean vector

$$\mu_J = \frac{\varphi}{\varphi + T} 1_N \mu_0 + \frac{T}{\varphi + T} \hat{\mu}, \quad [5]$$

and covariance matrix

$$\Lambda_J = \frac{\varphi}{\varphi + T} \Sigma. \quad [6]$$

The posterior mean [5] is an average of the prior mean and the sample mean vector. It turns out to be a simple and not a matrix-weighted average because the prior covariance matrix is specified as proportional to the asset covariance matrix in [4]. This linkage is also done by *Pastor* (2000), although it is hard to economically justify that correlations between estimation errors are equal to the asset correlations itself. Indeed, other Bayesian approaches for portfolio construction assume the prior covariance matrix to be diagonal, e.g., see *Black/Littermann* (1992) and *Herold* (2003).

The predictive pdf has the same mean as the posterior pdf and covariance matrix $\Sigma + \Lambda_J$. *Jorion* (1986) further demonstrates that φ can be estimated from the data:

$$\hat{\varphi} = \frac{N + 2}{(\hat{\mu} - \mu_0 1_N)' \Sigma^{-1} (\hat{\mu} - \mu_0 1_N)}, \quad [7]$$

where the denominator measures the observed dispersion of the sample means around the common mean.

As equation [5] shows, *Jorion* shrinks the sample means towards the MVP mean return. The longer the sample history, T , the weaker is the shrinkage. In the extreme, $T \rightarrow \infty$, the investor will use the sample means, $\mu_J = \hat{\mu}$, i.e. the estimator includes the sample mean as a special case. At the other extreme, with no uncertainty in the prior, $\varphi \rightarrow \infty$, the approach results in the MVP. In the (more interesting) cases in between these extremes, the MV efficient portfolio with risk aversion λ is shrunk towards the MVP. In fact, the investor with risk aversion λ moves to the left on the original MV efficient frontier and chooses a portfolio with a higher implicit lambda. *Frost/Savarino* (1986) provide an

extension by expanding the prior on the covariance matrix. Their prior is that assets have equal expected returns, equal variances and equal covariances. Hence they shrink towards the equally-weighted portfolio.

The prior that *Jorion* chooses is statistically motivated. *Stein* (1955) has shown that for multivariate normally distributed random variables, the sample mean is not an admissible estimator (for $N > 2$). *Stein* further proves that shrinkage estimators (which shrink the sample means towards a grand mean) have lower risk under a quadratic loss function. *Jorion* uses a methodologically different (Bayesian) approach, but his estimator also belongs to the class of shrinkage estimators. Hence, his approach is also called “Bayes/Stein estimation”. In addition to *Stein*, *Jorion* considers the impact of estimation risk on portfolio variance. Furthermore, he uses an empirical Bayes approach to obtain the parameters of the prior distribution. A drawback of the Bayes/Stein approach is that its prior assumption of equal expected returns is hard to be justified from an economic point of view. This assumption is removed by the approach we consider next.

3. Shrinking towards the market portfolio

In a series of articles, *Pastor* (2000), *Pastor/Stambaugh* (1999, 2000) and *Wang* (2001) have suggested to incorporate an asset-pricing model for asset allocation [hereafter called “the Pastor approach”].³ They combine the results of mean/variance optimization and the implications of an asset-pricing model, again using Bayesian inference. Their motivation is that mean/variance analysis, on the one side, only utilizes the data and bases portfolio selection on the first and second sample moments. It, however, completely ignores the potential usefulness of an asset-pricing model. Basing asset allocation only on an asset-pricing model, on the other side, makes no use at all of the time series of returns. In general, a model will neither be a perfect description of reality (by construction) nor will it be completely useless for decision making. To quote *Pastor* (2000, p. 179):

“By definition, every model is a simplification of reality. Hence, even if the data fail to reject the model, the decision maker may not necessarily want to use the model as a dogma. At the same time, the notion that models implied by finance theory could be entirely worthless seems rather extreme. Hence, even if the data reject the model, the decision maker may want to use the model at least to some degree.”

The approach developed by *Pastor* combines the sample information and the implications of an asset-pricing model. It is flexible to accommodate single- and multi-factor models. The implication of the CAPM is to hold the market portfolio, while an APT model implies to invest in a combination of the factor

³ For a review of these articles and an application to international diversification, see *Herold/Maurer* (2003).

portfolios. We consider only the CAPM. In this case, the prior expected excess returns are set equal to the expected excess returns implied by the CAPM, as given by

$$\mu_{\text{CAPM}} = \beta \mu_M, \quad [8]$$

where β is the $N \times 1$ beta vector and μ_M denotes the risk premium of the market portfolio. Hence, the expected excess returns are linked to the systematic risk in the prior. As in Jorion's approach, the posterior expected returns are a weighted average of the prior expected returns and the sample means:

$$\mu_P = \omega \mu_{\text{CAPM}} + (1 - \omega) \hat{\mu}, \quad [9]$$

where $\hat{\mu}$ now denotes the sample means of the excess returns (over the risk-free rate) and ω denotes the shrinkage factor. The sample means are shrunk towards the implied CAPM excess returns, hence the tangency portfolio (the MV efficient portfolio with the highest Sharpe Ratio) is shrunk towards the market portfolio. The shrinkage factor is a measure for the weight which is assigned to the CAPM and is given by

$$\omega = \frac{s^2 / \sigma^2}{s^2 / \sigma^2 + T / (1 + SR^2)}, \quad [10]$$

where SR is the Sharpe ratio of the market portfolio and s^2 is the average variance of the residual terms in the multivariate regression of the assets' excess returns on the excess returns of the market portfolio. σ measures the dispersion of the assets' alphas, i.e. the deviation of the assets' expected returns from the values implied by the CAPM. It is a measure of the investor's prior uncertainty in the CAPM. Thus, the degree of shrinkage depends on how much confidence the investor has into the validity of the CAPM and on the strength of the violations of the CAPM in the historical data. The optimal portfolio is approximately a linear average of the market portfolio and the tangency portfolio with a fraction of ω invested into the market portfolio. This holds only approximately because only the predictive mean is linear in ω but the covariance matrix is not.⁴

The prior precision, σ , can be either exogenously specified or by means of an empirical Bayes approach. Estimating σ from the data will lead to a shrinkage factor of 0.5, as Wang (2001) has shown, implying that the model and the data have equal impact on asset allocation. We report results only for this case, as our tests with different values for the shrinkage factors have given similar results.

⁴ The predictive mean is again equal to the posterior mean. For the expression of the predictive covariance matrix we refer to the original articles of Pastor (2000), Pastor/Stambaugh (1999, 2000) and Wang (2001).

V. BAYESIAN APPROACHES WHEN RETURNS ARE PREDICTABLE

1. Using Bayesian inference in predictive regressions

The Bayesian approaches summarized in the last section assume returns to be IID distributed. However, there is a multitude of articles that model expected stock returns as time-varying. Some examples are those by *Fama/Schwert* (1977), *Keim/Stambaugh* (1986), *Campbell* (1987), *Campbell/Shiller* (1988), *Fama/French* (1988, 1989), and *Ferson/Harvey* (1993). Time-varying expected returns do not automatically violate the efficient market hypothesis but might be explained by risk aversion changing over time; see *Fama* (1991) or *Cochrane* (1999).

If expected stock returns are time-varying, this also means that stock returns are partly predictable, and it should be possible to design investment strategies which exploit this predictability. Usually, the relationship between expected returns and instrumental variables is captured by linear regressions, called “predictive regressions”. The asset excess returns are regressed on (lagged) instrumental variables. Many studies that investigate the economic implications of return predictability within a MV portfolio optimization framework ignore estimation errors; see, e.g., the seminal work of *Solnik* (1993), as well as *Klemkosky/Bharati* (1995), and *Fletcher* (1997). However, the coefficients of the predictive regressions must be estimated, they are subject to estimation risk. Following contributions of *Beller et al.* (1998), *Connor* (1997) and *Herold/Maurer* (2004), we suggest to apply Bayesian multivariate regression models to account for estimation risk in the predictive regressions.

More formally, let $X = (1_T Z)$ be a $T \times (K + 1)$ matrix that consists of a vector of ones, and of the $T \times K$ matrix Z of the realizations of the instrumental variables. The instrumental variables are lagged by at least one period. The multivariate regression – regressing the asset excess returns on the instrumental variables – is given by

$$r = XB + U, \quad [11]$$

where r denotes the $T \times N$ matrix of excess returns and B denotes the $(K + 1) \times N$ matrix of regression coefficients. N is the number of assets, and K is the number of instrumental variables. B contains the N intercepts in the first row and the $K \times N$ slope coefficients in the other rows. The matrix U stores the residual terms which are assumed to be normally distributed:

$$\text{vec}(U) \sim N(0, \Sigma \otimes I_T), \quad [12]$$

where the $N \times N$ matrix Σ includes the variances and covariances of the residual terms and I_T is the $T \times T$ identity matrix.

The Maximum Likelihood estimator of the regression coefficients is given by

$$\hat{B} = (X'X)^{-1}X'r. \quad [13]$$

It is normally distributed as well:

$$\text{vec}(\hat{B}) \sim N(\text{vec}(B), (\Sigma \otimes (X'X)^{-1})). \quad [14]$$

Given the values of the instrumental variables at time T and the estimated regression coefficients, the expected excess return of the next period can be estimated:

$$E(r_{T+1} | \Phi_T) = \hat{B}'x'_T, \quad [15]$$

where Φ_T denotes the information set available at time T and the $1 \times (K+1)$ vector $x_T = (1 \ z_T)$ stores the actual values of the instrumental variables. Usually, these expected excess returns are interpreted as point forecasts and imported to an algorithm for portfolio optimization.

Information variables that have shown forecasting abilities in the literature are firm specific variables (e.g. the dividend yield or price-earnings-ratio)⁵ or macroeconomic variables that reflect the business cycle (e.g. interest rates, changes in interest rates, term and default spreads); see, among others, *Ang/Bekaert* (2003), *Chen et al.* (1986), *Fama* (1991), *Ferson/Harvey* (1991, 1993), *Harvey* (1995), *Hodrick* (1992). From an economic perspective, the former set of variables can be motivated by valuation models like the dividend discount model, while macroeconomic variables can be inferred from intertemporal asset pricing models. Furthermore, the lagged market return is often included to utilize the existence of autocorrelation effects (mean reversion or momentum).

To account for estimation risk or parameter uncertainty in the predictive regressions, we apply Bayesian regression models. Using Bayesian techniques in econometrics goes back to *Tiao/Zellner* (1964a, 1964b), *Zellner/Chetty* (1965), and *Zellner* (1971). In Bayesian regression models, a prior distribution of the regression coefficients and residual covariance matrix is specified and then combined with the likelihood function (i.e., the distribution of the sample estimates) to obtain the posterior distribution.

The natural-conjugate prior for B is given by

$$\text{vec}(B) | \Sigma \sim N(\text{vec}(B_0), \Sigma \otimes C), \quad [16]$$

where B_0 is the $(K+1) \times N$ matrix of the means and C is the $(K+1) \times (K+1)$ covariance matrix of the regression coefficients B . Σ is specified non-informative (i.e., the so-called Jeffrey's Prior). Then, the mean of the posterior distribution of the regression coefficients is a matrix-weighted mean of B_0 and \hat{B} :

$$\hat{B} = (C^{-1} + X'X)^{-1}(C^{-1}B_0 + X'X\hat{B}). \quad [17]$$

⁵ For recent discussions about the usefulness of dividend yields to predict risk premia, see *Goyan/Welsch* (2003) and the literature survey by *Rey* (2003).

2. Defining the shrinkage target

The crucial question is how to specify the parameters B_0 and C of the prior distribution in equation [16]. We have chosen the same two shrinkage targets as in the preceding section: the minimum-variance and the market portfolio. They serve as “neutral” allocation for the investor – if he has no forecasts or if the exogenous information is not reliable to him. To that end, the slope coefficients are set to zero in the prior, implying non time-varying, non-predictable returns and information variables having no impact on expected returns. The intercept is set differently depending on the shrinkage target: When shrinking to the MVP, the intercept is set to a common value across all regressions that is equal to the expected return of the MVP. When choosing the market portfolio as shrinking target, the intercepts are set to the implied market returns of the assets, given by [8].⁶ This way, we ensure that the optimal portfolio is shrunk towards the MVP in the first case and towards the market portfolio in the second case.

A possibility to specify the prior covariance matrix C is to apply the “Minnesota Prior”-technique. This procedure has been developed by the University of Minnesota and the Federal Reserve Bank of Minneapolis for Bayesian vector autoregressive (VAR) models in the 70s and 80s; see *Doan et al. (1984)*, *Litterman (1986)*, and *Todd (1984)*. The variances of the regression coefficients are controlled by a “tightness parameter”, θ , and scaled by the (inverse) variances of the instrumental variables:

$$C = \theta^2 \begin{pmatrix} \sigma_0^{-2} & 0 & \dots & 0 \\ 0 & \sigma_1^{-2} & & \\ \vdots & & \ddots & \\ 0 & & & \sigma_k^{-2} \end{pmatrix}. \quad [18]$$

C is a $(K + 1) \times (K + 1)$ diagonal matrix with $\sigma_0 = 1$. In the limit, $\theta \rightarrow \infty$, the values of the Bayesian regression coefficients equal those of the ML estimators. The prior specification in equation [18] is similar to *Zellner's (1986)* “g-prior”, where the variances of the regression coefficients are also controlled by a “hyper parameter”. In the “g prior” however, the coefficients are not assumed to be independent in the prior, but as correlated; the correlations are set equal to those in the sample. This implies that all slope coefficients are scaled downwards by the same magnitude (they are multiplied by a parameter g_0). In contrast, the prior in [18] shrinks the coefficients towards zero at different magnitudes.

⁶ A precise description of how to obtain the implied returns that lead to the market portfolio is given in *Kahn et al. (1996)*.

VI. EMPIRICAL OUT-OF-SAMPLE STUDY

1. Unconditional strategies

We use one of the longest data sets available, the U.S. stock market, to back-test the unconditional strategies, i.e. those strategies that follow the IID assumption and estimate the input parameters from return data only. The conditional strategies require a broader data set. Due to the limited availability of many of the instrumental variables, we use a shorter history of Euroland stock market indices in the next section.

We start with an investment universe of six size/book-to-market portfolios and then continue with an extended data set of 30 industry portfolios and 25 size/book-to-market portfolios.⁷ The sample period is from 1/1950 to 12/2003 whereby monthly excess returns are calculated using the 3month T-Bill rate as the risk-free rate of return. Based on this data set, we implement an “out-of-sample” back-testing procedure, which is a prominent approach in evaluating the impact of different investment strategies on portfolio performance under realistic conditions; see, among others, *Fletcher* (1997), *Jagannathan/Ma* (2003), *Klemkosky/Bharati* (1995), and *Solnik* (1993).

In each case, the out-of-sample period is from 1/1960 to 12/2003. We use a rolling window of length T to estimate the optimization input parameters. T is set to 30, 60, 90, and 120 months. E.g., for $T = 30$, portfolio weights are first based on the estimation period from 7/1957 to 12/1959. Using the excess returns of 1/1960, the first out-of-sample portfolio return can be calculated. Then the estimation period is rolled one month forward, and the next portfolio composition is based on 8/1957 to 1/1960. Regardless of sample size, T , this procedure results in a total of 528 non-overlapping monthly out-of-sample returns, which can also be considered as independent investment decisions with a holding period of one month.

At each point of time, a total of twelve portfolios is optimized: the tangency portfolios and optimal portfolios for risk aversions, λ , of 2 and 15, respectively, under the four approaches “classical MV optimization” (i.e., MV optimization based on sample estimates), “bootstrapping” (bootstrapping returns and calculating average weights over the optimal portfolios for the resampled returns),⁸ “Bayes/Stein” (shrinking towards the minimum-variance portfolio) and “Pastor” (shrinking to the market portfolio). The reason to not only consider the tangency portfolios, i.e. the portfolio maximizing the ratio of excess return to volatility (the risk adjusted performance measure proposed by Sharpe), is that in particular the Bayes/Stein strategy aims at (implicitly) increasing risk aversion and therefore it is desirable to control lambda explicitly. In our empirical

⁷ The data are taken from the website of *Kenneth French*. For a description of the construction of the portfolios, we refer to the web site: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

⁸ At each point of time, 500 bootstrapped return series were generated with a length equal to T .

study, it turns out that the implicit risk aversion of the tangency portfolio – that is, the value of λ in [1] that leads to the tangency portfolio – fluctuates widely over time.

These optimized portfolios using information on the expected returns are compared to each other and to strategies that refrain from estimating expected returns at all: the minimum-variance portfolio (MVP), the equally-weighted portfolio (EWP) and the U.S. market portfolio. The comparison of the different strategies is based on the empirical Sharpe ratio. Furthermore, monthly turnover is computed, and the weight structures are analyzed to obtain an indication which strategy leads to more stable weights over time and hence, lower transaction costs.

For the first investment universe of six size/book-to-market assets, Figure 1 presents the monthly mean excess returns, volatilities, and Sharpe ratios of the “naive” strategies EWP, market portfolio, and MVP and of the optimized strategies. Short selling is not permitted in each case. The Sharpe ratios are shown for the tangency portfolio as well as the two portfolios that are based on the mean/variance utility function. The latter ones could also be compared with respect to their “certainty-equivalent return” which is obtained by plugging in the out-of-sample mean and variance in [1]. The ranking of the strategies is materially not affected by either using the Sharpe ratio or the certainty-equivalent return.

The MVP succeeds in having the lowest volatility. Its Sharpe ratio is below that of the market and equally-weighted portfolio. The optimized strategies manage to produce a slightly higher Sharpe ratio in most cases. However, this increase is not statistically significant. Figure 1 reports the standard errors of the Sharpe ratio estimates based on the GMM procedure developed by *Lo* (2002). The standard errors are about one-third of the level of the Sharpe ratios.

Comparing the optimized strategies, it turns out that neither of the approaches that incorporate estimation risk can improve on the sample-based mean-variance portfolios. The volatility is somewhat reduced, but the Sharpe ratio is more or less the same. Increasing the rolling window length from 30 to 60 months can increase the Sharpe ratio. Using an even longer estimation window has no positive effect on performance. This could be an indication that the return distributions are not stationary, but we will investigate this issue in more detail in the next section.

When we repeat the analysis with short selling allowed, the volatilities sharply increase. The volatility for the optimized portfolios with high risk aversion of 15 is still acceptable, but the volatility of in particular the tangency portfolio explodes. The Sharpe ratios of the tangency portfolios deteriorate, often getting negative, while the two strategies based on the mean/variance utility function can increase the Sharpe ratio compared to the case without short-selling in Figure 1. Hence, the tangency portfolio does not accomplish its objective to maximize the Sharpe ratio in an out-of-sample context. But again, incorporating estimation risk cannot add value.

Finally, we implemented recursive estimation windows, starting with, e.g. 60 months, and then lengthening this window successively by one month when

FIGURE 1

OUT-OF-SAMPLE RESULTS FOR THE INVESTMENT UNIVERSE CONSISTING OF SIX SIZE/BOOK-TO-MARKET ASSET CLASSES

a) Mean excess returns, volatilities and Sharpe ratios of naive strategies

Portfolio	Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error
EWP	4.82	0.66	0.136	0.048
Market portfolio	4.25	0.64	0.150	0.044
MVP	3.96	0.50	0.126	0.052
$T = 30$	3.91	0.54	0.138	0.049
$T = 60$	4.15	0.52	0.126	0.051
$T = 90$	4.10	0.49	0.120	0.048
$T = 120$				

b) Mean excess returns, volatilities and Sharpe ratios of optimized strategies

T	Portfolio	Classical MV				Bootstrapping				Bayes/Stein				Pastor			
		Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error	Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error	Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error	Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error
30	TP	4.75	0.80	0.168	0.051	4.73	0.70	0.149	0.051	4.66	0.71	0.152	0.051	4.70	0.76	0.161	0.051
	$\lambda = 15$	4.41	0.65	0.147	0.051	4.36	0.67	0.154	0.051	4.23	0.54	0.128	0.051	4.07	0.50	0.122	0.050
	$\lambda = 2$	4.89	0.88	0.181	0.051	4.81	0.77	0.160	0.051	4.80	0.88	0.183	0.051	4.88	0.84	0.173	0.051
60	TP	4.69	0.84	0.178	0.051	4.65	0.73	0.156	0.051	4.55	0.82	0.179	0.051	4.61	0.80	0.174	0.050
	$\lambda = 15$	4.12	0.64	0.156	0.050	4.14	0.65	0.156	0.050	3.99	0.58	0.147	0.050	3.96	0.56	0.142	0.048
	$\lambda = 2$	4.90	0.82	0.167	0.051	4.72	0.76	0.161	0.050	4.80	0.79	0.166	0.050	4.92	0.76	0.155	0.051
90	TP	4.74	0.70	0.147	0.053	4.61	0.66	0.143	0.053	4.57	0.64	0.139	0.053	4.64	0.64	0.137	0.053
	$\lambda = 15$	4.24	0.55	0.129	0.053	4.23	0.56	0.133	0.053	4.17	0.52	0.124	0.052	4.15	0.51	0.123	0.051
	$\lambda = 2$	4.93	0.69	0.139	0.053	4.72	0.66	0.140	0.053	4.76	0.67	0.141	0.054	4.83	0.63	0.131	0.054
120	TP	4.72	0.76	0.161	0.051	4.49	0.73	0.161	0.051	4.52	0.66	0.146	0.050	4.57	0.70	0.154	0.050
	$\lambda = 15$	4.14	0.56	0.135	0.049	4.15	0.58	0.139	0.050	4.09	0.51	0.126	0.049	4.09	0.50	0.123	0.048
	$\lambda = 2$	4.92	0.85	0.173	0.052	4.71	0.79	0.168	0.052	4.78	0.81	0.169	0.052	4.83	0.78	0.161	0.052

The table displays the results of the out-of-sample strategies. Panel a) shows the (monthly) mean excess returns, standard deviations, Sharpe ratios, and GMM standard errors of the Sharpe ratios for the equally-weighted portfolio (EWP), the U.S. market portfolio, and the minimum-variance portfolio (MVP). Panel b) exhibits the statistics for the four strategies (classical MV, bootstrapping, Bayes/Stein, and Pastor) and for three risk aversions (tangency portfolio, $\lambda = 15$, $\lambda = 2$). The rolling window is varied from 30 to 60, 90, and 120 months. Short selling is not permitted.

FIGURE 2

TURNOVER AND WEIGHT STATISTICS FOR THE INVESTMENT UNIVERSE CONSISTING OF SIX SIZE/BOOK-TO-MARKET ASSET CLASSES

a) Average turnover in % (one-way)

T	Portfolio	Classical				MVP	EWP
		MV	Bootstrapping	Bayes/Stein	Pastor		
30	TP	14.37	10.43	13.58	12.73	6.90	0.77
	$\lambda = 15$	12.22	9.64	11.64	8.38		
	$\lambda = 2$	10.88	9.84	11.98	12.19		
60	TP	8.24	7.07	8.10	8.23	3.25	0.77
	$\lambda = 15$	6.81	6.13	4.90	4.02		
	$\lambda = 2$	7.48	7.07	7.89	8.80		
90	TP	5.78	5.22	5.42	5.87	2.33	0.77
	$\lambda = 15$	5.19	4.78	3.74	3.00		
	$\lambda = 2$	7.31	6.15	6.54	6.14		
120	TP	4.54	4.42	5.25	4.88	2.01	0.77
	$\lambda = 15$	3.72	4.08	2.86	2.44		
	$\lambda = 2$	5.69	4.96	5.12	4.30		

b) Weights of tangency portfolios (in %)

T	Asset	Classical MV		Bootstrapping		Bayes/Stein		Pastor	
		Mean weight	Standard deviation	Mean weight	Standard deviation	Mean weight	Standard deviation	Mean weight	Standard deviation
30	Small/Low	6.24	21.68	14.39	29.44	3.51	16.10	5.81	19.03
	Small/Mid	10.77	26.70	8.05	12.55	11.41	26.37	11.11	25.84
	Small/High	33.17	41.50	27.09	26.81	28.84	38.06	27.60	36.90
	Big/Low	19.62	37.67	14.59	25.57	17.94	34.89	21.43	36.01
	Big/Mid	13.47	29.17	15.58	21.23	19.21	33.30	18.20	29.84
	Big/High	16.73	32.60	20.31	21.46	19.09	32.45	15.86	29.48
60	Small/Low	1.25	6.94	8.48	21.17	0.26	3.50	1.44	7.10
	Small/Mid	3.27	14.23	6.18	8.07	5.50	17.64	4.97	17.07
	Small/High	46.50	44.46	34.95	29.25	41.11	42.37	39.72	40.79
	Big/Low	13.41	31.02	8.91	14.71	11.29	26.50	16.15	29.44
	Big/Mid	12.72	27.09	16.70	22.88	17.83	33.47	15.25	29.13
	Big/High	22.85	36.22	24.77	23.72	24.01	35.42	22.47	32.36
90	Small/Low	0.03	0.37	3.11	8.26	0.15	1.17	0.06	0.52
	Small/Mid	1.77	7.63	4.41	5.71	1.12	6.02	2.03	6.97
	Small/High	54.68	43.23	43.45	30.95	46.01	40.74	47.84	40.57
	Big/Low	6.77	17.64	5.75	8.64	4.49	11.70	10.62	19.31
	Big/Mid	15.27	30.85	19.01	26.07	21.18	36.27	18.45	31.39
	Big/High	21.49	34.02	24.27	23.86	27.05	35.17	21.01	30.11
120	Small/Low	0.03	0.59	0.76	2.11	0.38	2.39	0.07	0.69
	Small/Mid	0.65	3.23	3.40	4.09	0.39	2.13	0.77	2.99
	Small/High	54.99	42.89	45.48	29.87	43.98	39.84	48.20	39.39
	Big/Low	5.58	17.22	5.53	9.69	3.97	12.41	9.02	18.04
	Big/Mid	16.12	32.32	19.53	27.37	22.06	36.19	19.41	32.58
	Big/High	22.63	35.58	25.31	22.48	29.22	36.16	22.53	31.72

Panel a) displays the average monthly one-way turnover for the four strategies (classical MV, bootstrapping, Bayes/Stein, and Pastor) and for three risk aversions (tangency portfolio, $\lambda = 15$, $\lambda = 2$) as well as for the equally-weighted portfolio (EWP) and the minimum-variance portfolio (MVP). The rolling window is varied from 30 to 60, 90, and 120 months. Short selling is not permitted. Panel b) shows the means and standard deviations of the weights of the tangency portfolios.

rolling through time without discarding older information. This should diminish the impact of estimation risk but ignores that expected returns might be time-varying. It cannot improve the Sharpe ratio. In fact, after the first couple of out-of-sample months, the fluctuation in portfolio weights is low because the impact of new information is limited.

In practice, it is also important to implement strategies which produce low turnover. The three approaches that incorporate estimation risk can reduce turnover compared to the sample-based mean/variance case; see Panel a) of Figure 2. For small sample sizes, the tangency portfolio has the highest turnover, although its risk profile is between the two utility-based strategies. As expected, a larger estimation window implies a lower turnover. The weight structures of the tangency portfolios are displayed in the lower part of Figure 2. The mean weights and standard deviations (over time) are fairly similar across the different strategies.

It would be interesting to see whether incorporating estimation risk is more important in an investment universe with more assets. Shrinking to a target portfolio that has relatively constant weights (MVP or market portfolio) should at least reduce turnover more than if only few assets are included. This is why we have repeated the backtest for the case of 30 industry portfolios and 25 size/book-to-market portfolios. Figure 3 shows performance results and turnover for the non-optimized and optimized strategies. For space limitations, T is set to 60 months and we report only the results for the tangency portfolios without short selling.

The results are mixed. In the context of the industry portfolios, the optimized strategies lead to Sharpe ratios below that of the market and minimum-variance portfolio, while with size/book-to-market assets, they can increase the Sharpe ratio. In neither case, the differences are statistically significant, as the standard errors are again fairly large. The reason for the disappointing performance of the approaches that incorporate estimation risk might be that the returns are not IID distributed. This motivates for the inclusion of exogenous information. In the next section, we replace the sample means by conditional expected return estimates. Finally, we note that in particular the bootstrap approach can reduce turnover, albeit the turnover is still considerably larger than that of the MVP.

2. Incorporating exogenous information

The objective of the analysis in this section is twofold: First, to check the robustness of the results we repeat the analysis for the unconditional strategies (i.e. those based on the IID assumption of the return generation process) from above using another investment universe. Second, to assess the importance of incorporating exogenous information, we investigate conditional investment strategies where the expected returns for the next period are derived from the predictive regressions.

FIGURE 3

OUT-OF-SAMPLE RESULTS FOR THE INVESTMENT UNIVERSES CONSISTING OF 30 INDUSTRY PORTFOLIOS
AND 25 SIZE/BOOK-TO-MARKET ASSET CLASSES

Performance figures and turnover of naive and tangency portfolios for $T = 60$

a) Investment universe: 30 industry portfolios

Portfolio	Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error	Turnover (in %)
<i>Naive strategies</i>					
EWP	4.69	0.55	0.118	0.047	1.33
Market portfolio	4.26	0.63	0.147	0.044	0.00
MVP	3.49	0.49	0.141	0.049	6.34
<i>Tangency portfolios</i>					
Classical MV	4.93	0.48	0.097	0.048	14.97
Bootstrapping	4.71	0.54	0.114	0.049	13.46
Bayes/Stein	4.61	0.45	0.098	0.048	14.29
Pastor	4.58	0.50	0.109	0.047	12.89

b) Investment universe: 25 size/book-to-market portfolios

Portfolio	Volatility (in %)	Mean excess return (in %)	Sharpe Ratio	GMM std. error	Turnover (in %)
<i>Naive strategies</i>					
EWP	5.03	0.69	0.138	0.049	0.88
Market portfolio	4.25	0.64	0.150	0.044	0.00
MVP	3.93	0.53	0.134	0.050	6.66
<i>Tangency portfolios</i>					
Classical MV	4.84	0.85	0.176	0.052	16.98
Bootstrapping	4.91	0.79	0.162	0.054	10.68
Bayes/Stein	4.73	0.81	0.172	0.052	17.01
Pastor	4.59	0.78	0.169	0.050	15.02

The table displays the (monthly) mean excess returns, standard deviations, Sharpe ratios, GMM standard errors of the Sharpe ratios, and monthly turnovers for the naive and optimized portfolio strategies. The rolling window is set to 60 months. Short selling is not permitted.

The investment universe comprises the 10 Datastream sector indices for Euroland. The sample consists of monthly data from 12/88 to 6/04 (187 observations).⁹ The performance indices are converted into excess returns, with the 3-month FIBOR proxying the risk-free rate. The instrumental variables comprise 10 macroeconomic variables and valuation ratios: the term spread between

⁹ The Datastream indices start 12/73. The starting point 12/88 is due to the instrumental variables, which were partly not available before.

long-term bonds and the 3-month rate, the short-term spread between the 3-month rate and 1-month rate, the default spread, oil price, industrial production, Euroland inflation (CPI), OECD leading indicator, price-earnings-ratio (PER), dividend yield, and lagged market excess return. The default spread is defined as the yield differential between BBB- and AAA-rated bonds. Due to longer data availability, the US default spread is used. Market return, PER, and dividend yield refer to the Datastream EMU index, which is made up of the 10 sector indices on a value-weighted basis. All variables are market-wide, none are industry-specific. Industrial production, inflation, and OECD leading indicator are lagged by 3 months to take account of the publication lag. All other variables are lagged by 1 month.

The expected returns for the next period that enter the conditional strategies are derived from the predictive regressions. At the beginning of each month, the regression model is re-estimated, based on data available at that point in time. Using the actual values of the instrumental variables, expected returns are predicted. The estimation period to compute the regression coefficients and the remaining input parameters is initially set to 30 months. The first estimation period is from 12/88 to 6/91. Then, excess returns for 7/91 are predicted. The covariance matrix is also based on the period from 12/88 to 6/91. Then, the estimation period is successively rolled one month forward, leading to 156 out-of-sample returns for each strategy. The procedure is repeated with an estimation window of 60 months. The out-of-sample period is then from 1/94-6/04 (126 months).

Figure 4 summarizes the performance and turnover statistics. The unconditional strategies are based on historical (excess) returns only. The conditional strategies use the output from the predictive regressions. In the “raw forecasts” strategy, the conditional expected returns are inputted to the mean/variance optimizer, while in the two Bayesian strategies, they go through a refinement process and are shrunk towards the expected returns implied by the minimum-variance (“Bayes_MVP”) or market portfolio (“Bayes_MP”). Following *Doan et al.* (1984), the tightness parameter, θ , is set to 0.2. In each case, the objective function is to maximize the Sharpe ratio.

For all investment strategies, the sample covariance matrix, based on the last T observations, is used. Since portfolio weights are much more sensitive to changes in expected returns compared to risk parameters, and since we want to evaluate different estimators for expected returns, it seems reasonable to use the same covariance matrix for all strategies. Short selling is not permitted.

Consider first an estimation window of 30 months (Panel a in Figure 4). The Sharpe ratio of the market portfolio is 0.10. The tangency portfolio as well as the two unconditional Bayesian approaches (Bayes/Stein and Pastor) can improve the Sharpe ratio a little, but the increase is again small. The minimum-variance portfolio turns out to be best among the unconditional strategies. Incorporating exogenous information and performing the predictive regressions, the Sharpe ratio substantially increases. It takes on a value of more than 0.20 in the strategy based on the raw forecasts, almost twice as that of the tangency

FIGURE 4

OUT-OF-SAMPLE RESULTS FOR THE INVESTMENT UNIVERSE CONSISTING OF 10 EUROPEAN SECTOR INDICES

a) Results for estimation window $T = 30$

Strategy	Mean (in %)	Volatility (in %)	Sharpe ratio	GMM std. error	% negative returns	Turnover (in %)
<i>Unconditional strategies</i>						
Minimum-variance	0.63	3.78	0.168	0.093	39.1%	6.19
Tangency portfolio	0.66	5.65	0.116	0.093	46.8%	22.24
Equally-weighted	0.65	4.98	0.131	0.093	42.9%	1.25
Market Portfolio	0.51	5.10	0.100	0.093	42.9%	0.44
Bayes/Stein	0.60	5.23	0.116	0.086	48.1%	19.67
Pastor	0.58	5.13	0.114	0.094	46.8%	11.39
<i>Conditional strategies</i>						
Raw forecasts	1.36	6.40	0.212	0.083	39.1%	62.40
Bayes_MVP	1.19	6.55	0.182	0.079	36.5%	52.37
Bayes_MP	1.19	6.55	0.182	0.079	36.5%	52.38

b) Results for estimation window $T = 60$

Strategy	Mean (in %)	Volatility (in %)	Sharpe ratio	GMM std. error	% negative returns	Turnover (in %)
<i>Unconditional strategies</i>						
Minimum-variance	0.77	3.88	0.198	0.106	34.9%	4.18
Tangency portfolio	1.05	6.00	0.176	0.105	48.4%	14.99
Equally-weighted	0.71	5.23	0.136	0.102	42.1%	1.35
Market Portfolio	0.57	5.37	0.105	0.103	42.1%	0.48
Bayes/Stein	1.11	5.45	0.203	0.098	44.4%	15.87
Pastor	0.81	5.46	0.148	0.107	44.4%	7.79
<i>Conditional strategies</i>						
Raw forecasts	1.39	7.05	0.198	0.093	36.5%	49.87
Bayes_MVP	1.49	6.75	0.222	0.097	37.3%	45.40
Bayes_MP	1.50	6.75	0.223	0.097	37.3%	45.69

Mean, volatility, and Sharpe ratio are monthly statistics.

“% negative returns” denotes the fraction of months where the excess return is negative.

“Turnover” is the monthly one-way turnover in percentage terms.

portfolio based on sample means. Due to the large GMM standard errors, this increase is not statistically significant, however it definitely is economically:¹⁰ The average monthly excess return rises from 66 bp to 136 bp, or in annual terms, the investor earns 840 bp more on the average in each year. The standard

¹⁰ The GMM standard errors of the Sharpe ratio estimates are higher than in Figures 1-3 because the number of out-of-sample periods is much smaller.

deviation moderately increases from 5.65% to 6.40%. Also, the number of months with negative excess returns is reduced.

The optimal portfolio based on the raw forecasts shows a quite large turnover. Taking account of estimation risk in the parameters of the predictive regressions leads to a slightly lower Sharpe ratio and reduces turnover. If we had maximized a mean-variance utility function with a moderate risk aversion (instead of maximizing the Sharpe ratio) the turnover would have been much less.

Using a larger estimation window of 60 months, the results are qualitatively similar albeit the differences are less pronounced (see Panel b). The Sharpe ratios can be improved by substituting sample means with forecasts from the predictive regressions. The two Bayesian conditional approaches now show the highest Sharpe ratio. The choice of the shrinkage target does not seem to be important because both Bayesian investment strategies lead to a very similar risk/return profile. The reason is that the only difference is the different intercept in the prior of the regression coefficients, B_0 . The slope coefficients of the predictive regressions are shrunk towards zero with the same magnitude under both alternatives. Choosing a smaller tightness parameter would lead to substantial different performance statistics, but then the shrinkage target would have a large impact at the same time.

VII. CONCLUSION

Large estimation errors in real-life financial data make MV optimization hard to apply in practice. Several approaches have been developed under the IID setting to incorporate estimation risk into portfolio choice. The heuristic approaches aim at producing more diversified portfolios either by introducing constraints on portfolio weights or by averaging over portfolio weights that have been obtained through bootstrapping the return data. In contrast, the Bayesian approaches directly adjust the inputs. It can be shown that imposing constraints is equivalent to using a modified set of optimization inputs. Therefore, from a theoretical standpoint, it is preferable to tackle the problem directly and use the Bayesian approaches that explicitly adjust the inputs than to implicitly change them through constraints. Furthermore, the Bayesian approaches have a strong foundation in statistical decision theory.

In this article, we transfer those Bayesian approaches that operate under the IID assumption to the setting where returns are partly predictable. We employ Bayesian regression models to account for the estimation risk attached to the coefficients in predictive regressions. In an out-of-sample study, we compare the performance of these approaches and find a couple of interesting results. First short-selling restrictions prove to be absolutely necessary to reduce portfolio volatility and turnover. Second all of the approaches assuming IID returns cannot improve on MV efficient portfolios. At least for the assets and time period studied, they lead to portfolios which are fairly similar to MV optimized portfolios in terms of the Sharpe ratio. We are not able to find them to be

systematically superior to the market portfolio or minimum-variance portfolio which refrains from estimating expected returns at all.

The disappointing performance of even “sophisticated” strategies like the Bayesian approaches indicates that historical returns are not very useful for estimating future expected returns. The IID assumption seems to be violated in the data. Indeed, the conditional strategies that use instrumental variables to predict stock returns can substantially improve the out-of-sample performance. Although the increase in the Sharpe ratio is not statistically significant, it is economically. It should also be possible to further enhance the performance of the conditional strategies by choosing different instrumental variables (e.g., industry-specific ones instead of market-wide variables) or by using non-linear forecasting models. In any case, the approaches that incorporate estimation risk reduce volatility and turnover. It seems to be advantageous to employ Bayesian models for portfolio construction.

REFERENCES

- ANG, A. and BEKAERT, G. (2003) Stock return predictability: Is it there?, *Working Paper Columbia University*.
- BANZ, R.W. (1997) Editorial: Zero correlation between theory and practice in financial economics?, *Finanzmarkt und Portfoliomanagement* **11**, 383-391.
- BARRY, C.B. (1974) Portfolio analysis under uncertain means, variances, and covariances, *Journal of Finance* **29**, 515-522.
- BELLER, K.R., KLING, J.L. and LEVINSON, M.J. (1998) Are industry stock returns predictable?, *Financial Analysts Journal* **54**, September/October, 42-57.
- BEST, M. and GRAUER, R. (1991) On the sensitivity of mean variance efficient portfolios to changes in asset means. Some analytical and computational results, *The Review of Financial Studies* **4**, No. 2, 315-342.
- BLACK, F. and LITTERMAN, R. (1992) Global Portfolio Optimization, *Financial Analysts Journal*, September/October, 8-43.
- CAMPBELL, J.Y. (1987) Stock returns and the term structure, *Journal of Financial Economics* **18**, 373-399.
- CAMPBELL, J.Y. and SHILLER, R.J. (1988) Stock prices, earnings, and expected dividends, *Journal of Finance* **43**, 661-676.
- CAMPBELL, J.Y. and VICEIRA, L.M. (2002) Strategic asset allocation, Clarendon Lectures in Economics: *Oxford University Press*.
- CHEN, N., ROLL, R. and ROSS, S.A. (1986) Economic forces and the stock market, *Journal of Business* **59**, 383-403.
- CHOPRA, V. and ZIEMBA, W.T. (1993) The effect of errors in means, variances, and covariances on optimal portfolio choice, *Journal of Portfolio Management* **19**, 6-11.
- CHOPRA, V.K., HENSEL, C.R. and TURNER, A.L. (1993) Massaging mean/variance inputs: Returns from alternative global investment strategies in the 1980s, *Management Science* **39**, 845-855.
- COCHRANE, J.H. (1999) Portfolio advice for a multifactor world, in: *Federal Reserve Bank of Chicago Economic Perspectives* **23**, No. 3, 59-78
- CONNOR, G. (1997) Sensible return forecasting for portfolio management, in: *Financial Analysts Journal* **53**, September/October, 44-51.
- DEMIGUEL, V., GARLAPPI, L. and UPPAL, R. (2005) How inefficient is the 1/N asset allocation strategy?, *Working Paper*.
- DOAN, T., LITTERMAN, R. and SIMS, C. (1984) Forecasting and conditional projecting using realistic prior distributions, *Econometric Reviews* **3**, 1-100.
- EFRON, B. (1979) Bootstrap methods: Another look at the jackknife, *Annals of Statistics* **7**, 1-26.
- EICHHORN, D., GUPTA, F. and STUBBS, E. (1998) Using constraints to improve the robustness of asset allocation, *Journal of Portfolio Management* **24**, Spring, 41-48.
- FAMA, E.F. (1991) Efficient capital markets: II, *Journal of Finance* **46**, 1575-1617.

- FAMA, E.F. and FRENCH, K. (1988) Dividend yields and expected stock returns, *Journal of Financial Economics* **22**, 3-25.
- FAMA, E.F. and FRENCH, K. (1989) Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* **25**, 23-49.
- FAMA, E.F. and SCHWERT, G.W. (1977) Asset returns and inflation, *Journal of Financial Economics* **5**, 15-146.
- FERSON, W.E. and HARVEY, C.R. (1991) The variation of economic risk premiums, *Journal of Political Economy* **99**, 385-415.
- FERSON, W.E. and HARVEY, C.R. (1993) The risk and predictability of international equity returns, *Review of Financial Studies* **6**, 527-566.
- FLETCHER, J. (1997) An investigation of alternative estimators of expected returns in mean-variance analysis, *Journal of Financial Research* **20**, 129-143.
- FLETCHER, J. and HILLIER, J. (2001) An examination of resampled portfolio efficiency, *Financial Analysts Journal* **57**, September/October, 66-74.
- FROST, P. and SAVARINO, J. (1986) An empirical Bayes approach to efficient portfolio selection, *Journal of Financial and Quantitative Analysis* **21**, 293-305.
- FROST, P. and SAVARINO, J. (1988) For better performance: Constrain portfolio weights, *Journal of Portfolio Management* **14**, 29-34.
- GOYAN, A. and WELSCH, I. (2003) Predicting the equity premium with dividend ratios, *Management Science* **49**, 639-654.
- GRAUER, R. and HAKANSSON, N. (1995) Stein and CAPM estimators of the means in asset allocation, *International Review of Financial Analysis* **4**, 35-66.
- GRAUER, R. and SHEN, F. (2000) Do constraints improve portfolio performance?, *Journal of Banking and Finance* **24**, 1253-1274.
- GREEN, R. and HOLLIFIELD, B. (1992) When will mean-variance efficient portfolios be well diversified?, *Journal of Finance* **47**, 1785-1809.
- GRINOLD, R.C. and EASTON, K.K. (1998) Attribution of performance and holdings, in: Mulvey, J.M./Ziemba, W.T. (ed.) *Worldwide asset and liability modeling*, Cambridge, 87-113.
- HARVEY, C.R. (1995) Predictable risk and returns in emerging markets, *Review of Financial Studies* **8**, 773-816.
- HEROLD, U. (2003) Portfolio construction with qualitative inputs, *Journal of Portfolio Management* **29**, 61-72.
- HEROLD, U. and MAURER, R. (2003) Bayesian asset allocation and U.S. domestic bias, *Financial Analysts Journal* **59**, November/December, 54-65.
- HEROLD, U. and MAURER, R. (2004) Tactical asset allocation and estimation risk, *Financial Markets and Portfolio Management* **18**, 39-57.
- HODRICK, R.J. (1992) Dividend yields and expected stock returns, *Review of Financial Studies* **3**, 357-386.
- JAGANNATHAN, R. and MA, T. (2003) Risk reduction in large portfolios: Why imposing the wrong constraints helps, *Journal of Finance* **58**, 1651-1683.
- JOBSON, J.D. and KORKIE, B.M. (1981a) Performance hypothesis testing with the Sharpe and Treynor measures, *Journal of Finance* **36**, 888-908.
- JOBSON, J.D. and KORKIE, B.M. (1981b) Putting Markowitz theory to work, *Journal of Portfolio Management* **7**, 70-74.
- JORION, P. (1985) International portfolio diversification with estimation risk, *Journal of Business* **58**, 259-278.
- JORION, P. (1986) Bayes-Stein estimation for portfolio analysis, *Journal of Financial and Quantitative Analysis* **21**, 279-292.
- JORION, P. (1991) Bayesian and CAPM estimators of the mean: Implications for portfolio selection, *Journal of Banking and Finance* **15**, 717-727.
- JORION, P. (1992) Portfolio optimization in practice, *Financial Analysts Journal* January/February **48**, 68-74.
- KAHN, R.N., ROULET, J. and TAJBAKSH, S. (1996) Three steps to global asset allocation, *Journal of Portfolio Management* **22**, 23-31.
- KEIM, D. and STAMBAUGH, R. (1986) Predicting returns in the stock and bond markets, *Journal of Financial Economics* **17**, 357-390.
- KLEIN, R. and BAWA, V. (1976) The effect of estimation risk on optimal portfolio choice, in: *Journal of Financial Economics* **3**, 215-231.

- KLEMKOSKY, R. and BHARATI, R. (1995) Time-varying expected returns and asset allocation, in: *Journal of Portfolio Management* **21**, Summer, 80-87.
- LITTERMAN, R. (1986) Forecasting with Bayesian vector autoregressions – Five years of experience, *Journal of Business and Economics Statistics* **4**, 25-38.
- LO, A. (2002) The statistics of Sharpe ratio, *Financial Analysts Journal* **58**, July/August, 36-52.
- MARKOWITZ, H. (1959) *Portfolio selection: Efficient diversification of investments*, Cambridge.
- MICHAUD, R.O. (1989) The Markowitz optimization enigma: Is 'optimized' optimal?, *Financial Analysts Journal* **45**, January/February, 31-42.
- MICHAUD, R.O. (1998) Efficient asset management, *Harvard Business School Press*, Boston.
- PASTOR, L. (2000) Portfolio selection and asset pricing models, in: *Journal of Finance* **55**, 179-223.
- PASTOR, L. and STAMBAUGH, R.F. (1999) Costs of equity capital and model mispricing, in: *Journal of Finance* **54**, 67-121.
- PASTOR, L. and STAMBAUGH, R.F. (2000) Comparing asset pricing models: An investment perspective, *Journal of Financial Economics* **56**, 335-381.
- REY, D. (2003) Stock market predictability: Is it there? A critical review, University of Basel, *WWZ Working Paper* No. 12-03.
- SCHERER, B. (2002) Portfolio resampling: Review and critique, *Financial Analysts Journal* **58**, November/December, 98-109.
- SOLNIK, B. (1993) The performance of international asset allocation strategies using conditioning information, *Journal of Empirical Finance* **1**, 33-55.
- STEIN, C. (1955) Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proceeding of the 3rd Berkely Symposium on Probability and Statistics* 1, Berkely: University of California Press.
- TIAO, G.C. and ZELLNER, A. (1964a) Bayes' Theorem and the use of prior knowledge in regression analysis, *Biometrika* **51**, 219-230.
- TIAO, G.C. and ZELLNER, A. (1964b) On the Bayesian estimation of multivariate regression, *Journal of the Royal Statistical Society* **26**, Series B, N277-285.
- TODD, R. (1984) Improving economic forecasting with Bayesian vector autoregression, *Federal Reserve Bank of Minneapolis Quarterly Review* **8**, 18-29.
- WANG, Z. (2001) A shrinkage approach to asset allocation using asset-pricing models, Working Paper Columbia University.
- ZELLNER, A. (1971) *An introduction to Bayesian inference in econometrics*, New York: Wiley.
- ZELLNER, A. (1986) On assessing prior distributions and Bayesian regression analysis using g-prior distributions, in: Goel, P./Zellner, A. (ed.) *Bayesian inference and decision techniques – Essays in honour of Bruno de Finetti*, North Holland, Chapter 15, 233-243.
- ZELLNER, A. and CHETTY, K. (1965) Prediction and decision problems in regression models from the Bayesian point of view, *Journal of the American Statistical Association* **60**, 608-616.

DR. ULF HEROLD

*Metzler Investment GmbH and Johann Wolfgang Goethe University,
Frankfurt/Main, Germany
Grosse Gallusstr. 18, 60311 Frankfurt/Main, Germany
Tel.: +49 69 2104 1138
Email: UHerold@Metzler.com*

PROF. DR. RAIMOND MAURER

*Johann Wolfgang Goethe University, Frankfurt/Main, Germany
Faculty of Business and Economics
Department of Finance
Senckenberganlage 31-33 (Uni-PF 58), 60054 Frankfurt/Main, Germany
Tel.: +49 69 798 25227
Email: RMaurer@wiwi.uni-frankfurt.de*